ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2016, 52, 3684

Received 15th December 2015, Accepted 26th January 2016

DOI: 10.1039/c5cc10317g

www.rsc.org/chemcomm

Synthesis of activated 3′-amino-3′-deoxy-2-thio-thymidine, a superior substrate for the nonenzymatic copying of nucleic acid templates†

Enver Cagri Izgu,[ab] Seung Soo Oh[ab] and Jack W. Szostak*[abc]

We present a scalable synthesis of 3′-amino-3′-deoxy-2-thio-thymidine-5′-phosphoro-2-methylimidazolide, an activated monomer that can copy adenosine residues in nucleic acid templates rapidly without a polymerase. The sulfur atom substitution enhances the rate of template copying by 5-fold compared with the 3′-amino-3′-deoxy-T monomer, while the 3′-amino monomers exhibit a 2- to 30-fold enhancement compared with their ribonucleotide counterparts.

Nonenzymatic template-directed replication of nucleic acids has been hypothesized to be the mechanism of information transfer in primitive cells prior to the advent of ribozyme polymerases.[1] Early efforts involving high-energy nucleotide monomers such as 5′-phosphoro-2-methylimidazolides (or 2-MeImpNs) (Fig. 1, top) showed that RNA templates consisting of C residues can be copied by 2-MeImpG in hours to days in the presence of divalent cations (typically Mg^{2+}).[2] However, no enzyme-free process has yet been discovered that enables the rapid and efficient copying of mixed-sequence RNA templates with activated ribonucleotide monomers. This problem has stimulated interest in alternative nucleic acids that might exhibit faster replication chemistry; such polymers are of interest both with respect to the origin of life and in the context of designing artificial life forms based on non-biological chemistry. The most promising non-biological nucleic acids are the phosphoramidate polymers, which are assembled from nucleotides with an amino group on the sugar instead of the less nucleophilic hydroxyl. N3′-P5′-linked phosphoramidate DNA[3,4] (3′-NP-DNA, Fig. 1, bottom) stands out as an attractive alternative genetic model because it adopts a helical geometry that is similar to that of A-form RNA.[5] A-form geometry is the preferred conformation for the nonenzymatic template-directed oligomerization of activated ribonucleotides[6,7] most likely because the A-form double helix of RNA brings the 3′-OH group of the primer in line with the leaving group of the incoming monomer. We have previously shown[8] that activated 3′-amino-2′,3′-dideoxynucleotide monophosphates (3′-NH_2-2-MeImpddNs) (Fig. 1, bottom) rapidly polymerize on short, homopolymeric DNA, RNA, and locked nucleic acid (LNA) templates. We also found[9] that replacing 3′-amino-T with 3′-amino-2-thio-T enhances the rate and fidelity of 3′-NP-DNA synthesis in the copying of DNA, RNA and 3′-NP-DNA templates. However, further progress in the

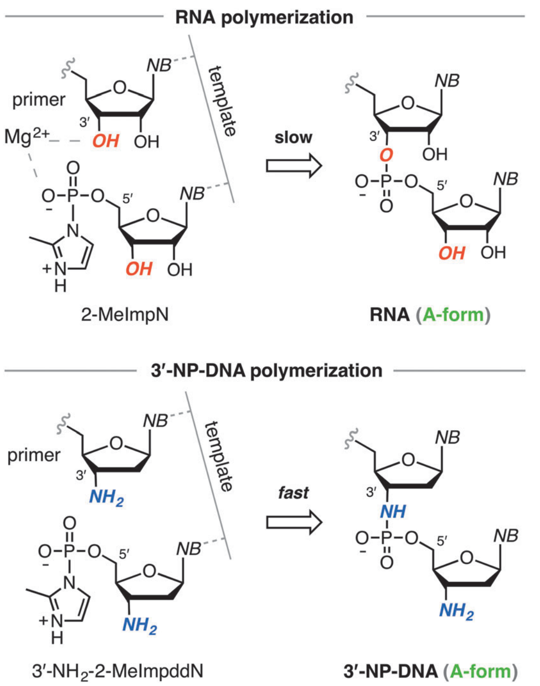


Fig. 1 Template-directed polymerization of RNA (top) and 3-NP-DNA (bottom). *NB* denotes nucleobase.

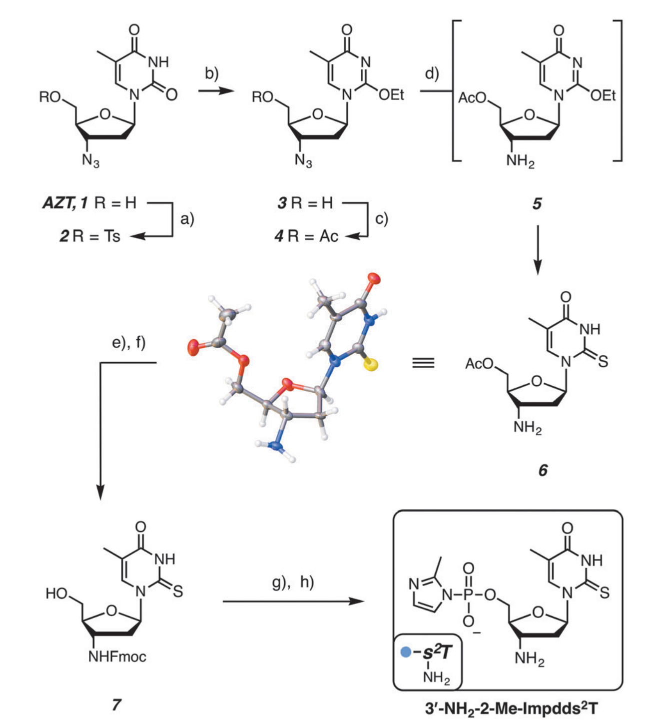
[a] Howard Hughes Medical Institute, Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA. E-mail: szostak@molbio.mgh.harvard.edu
[b] Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA
[c] Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford St., Cambridge, Massachusetts 02138, USA

† Electronic supplementary information (ESI) available. CCDC 1440493. For ESI and crystallographic data in CIF or other electronic format see DOI: 10.1039/c5cc10317g

development and quantitative analysis of this system has been hindered by lack of access to the critical substrate, the 5′-phosphoro-imidazolide of 3′-amino-2-thio-T (3′-NH_2-2-MeImp-dds^{2}T), which could previously be synthesized only in small quantities because of the expensive starting material and the very low yield of the desired product.

Here, we present a concise and scalable synthesis of 3′-amino-2-thio-T and the corresponding 5′-phosphoro-imidazolide, 3′-NH_2-2-MeImpdds^{2}T. We also present the first quantitative investigation of template-copying kinetics using this activated nucleotide. Our results show that nonenzymatic 3′-NP-DNA synthesis using the 2-thio modified 3′-amino-T phosphoro-imidazolide monomer is significantly faster than with unmodified 3′-amino-T, and is also considerably faster than RNA synthesis with activated U and 2-thio-U.

Our earlier attempts to establish a concise synthetic route to 3′-NH_2-2-MeImpdds^{2}T involved regioselective thio-carbonylation[10] of 4-*O*-protected 3′-azido-3′-deoxy-thymidine. However, these substrates afforded only deglycosylation products using Lawesson's reagent, an observation later reported[11] for 4-*O*-mesityl-3′,5′-*O*-di(TBS)-thymidine. We then studied nucleophilic ring opening of 2-thio-2,3′-cyclonucleosides by azides at the 3′-position, but observed little to no conversion (by ^{1}H NMR analysis) of the substrates in the presence of either excess LiN_3 or $TMSN_3$ accompanied by various Lewis acids [*e.g.*, $Hg(OAc)_2$, $Er(OTf)_3$ and $Yb(O\text{-}iPr)_3$].



Scheme 1 Synthesis of 3′-NH_2-2-MeImpdds^{2}T. Reaction conditions: (a) TsCl, pyr, 0 to 20 °C, 6 h, 73%; (b) DBU, EtOH, 85 °C, 24 h, 81%; (c) DMAP (cat.), Ac_2O, 20 °C, 12 h, >95%; (d) H_2S (gas), TMG, pyr, 0 to 20 °C, 16 h, 53%; (e) 7N NH_3 in MeOH, 20 °C, 92%; (f) Fmoc-OSu, Na_2CO_3 (aq), pyr, 0 to 20 °C, 6 h, 78%; (g) 1. $POCl_3$, 2,6-lutidine, $PO(OMe)_3$, 3 Å MS, 0 to 20 °C, 2 h; 2. 2-Me-imidazole, 2 h, 30%; (h) piperidine, DMF, 0 °C, 0.5 h, 85%. X-ray structure: H: white, C: gray, O: red, N: blue, and S: yellow.

Our successful strategy to access 3′-NH_2-2-MeImpdds^{2}T (Scheme 1) commenced with the 5′-tosylation[12] of 3′-azido-3′-deoxythymidine (AZT) **1**.[13] Intramolecular displacement of the 5′-tosylate by the C2-oxyanion formed in the presence of a Brønsted base (*e.g.*, DBU) yielded the 2,5′-*O*-anhydro-cyclo-nucleoside, and subsequent ring-opening in refluxing ethanol afforded the 2-ethoxythymidine **3** in 59% overall yield from **1**. After converting **3** into the acetate **4**, we were able to incorporate the sulfur atom into the nucleobase using H_2S in the presence of tetramethylguanidine (TMG)[14] (see Fig. S1 in the ESI† for details of the reaction setup). ^{1}H NMR analysis of an aliquot of the crude reaction mixture after 1 hour revealed that **4** was fully consumed, while two new species were formed: 3′-amino-2-ethoxythymidine **5** and 3′-amino-2-thio-thymidine **6**, in a molar ratio of 2 : 1 (**5**/**6**). The relative abundance of **6** continued to increase as the reaction progressed [up to *ca.* 1 : 3 (**5**/**6**) after 6 hours]. We did not observe (by either ESI or ^{1}H NMR analysis) the formation of any 3′-azido-2-thio-thymidine, suggesting that the incorporation of sulfur was slower than the reduction of the 3′-azide, and that **6** was likely formed from **5**. The structure of the 2-thio nucleoside **6** was confirmed by both ^{1}H–^{1}H gCOSY NMR spectroscopy and X-ray crystallography (see the ESI†). We then converted **6** into the phosphoroimidazolide precursor **7** *via* 5′-deacetylation and Fmoc protection of the 3′-amine. A one-flask 5′-*O*-phosphorylation and 2-methyl-imidazolide synthesis, followed by the removal of Fmoc, provided 3′-NH_2-2-MeImpdds^{2}T in 26% overall yield from **7**.

With 3′-NH_2-2-MeImpdds^{2}T in hand we proceeded to carry out nonenzymatic primer extension experiments to quantitatively interrogate the effect of 2-thio substitution on RNA template-copying rates (Fig. 2). We used a primer/template complex composed of a DNA primer strand ending in a 3′-NH_2-G and a complementary RNA template strand containing a 5′-C_2A_4 overhang. pK_a of the protonated 3′-amine of 3′-amino-2′,3′-dideoxy-2-thio-thymidine is 7.5 (see the ESI†), similar to that reported for 3′-amino-2′,3′-dideoxy-T (7.7).[15] 3′-NH_2-2-MeImpddNs tends to undergo intra-molecular cyclization due to the proximity of the primary 3′-amine group to the phosphorus electrophile.[8] Because the half-life ($t_{1/2}$) of 3′-NH_2-2-MeImpddT is 1.2 h[8] and that of 3′-NH_2-2-MeImpdds^{2}T is 1.3 h (see the ESI† for details) under optimized primer extension conditions [100 mM 1-(2-hydroxyethyl)-imidazole, pH 7.5, 4 °C], we tracked primer extension only up to a maximum of 1 h. We determined observed rate constants k_{obs} for the first step of the primer extension by following the loss of unreacted primer over time (Fig. 2).

At a 10 mM initial concentration of 3′-NH_2-2-MeImpddT (Fig. 2, left), k_{obs} of primer extension was 0.42 h^{-1} (Table 1, entry 1). Notably, 2-thio modification led to about a 5-fold rate enhancement (Fig. 2, right), such that k_{obs} for 10 mM 3′-NH_2-2-MeImpdds^{2}T was 1.92 h^{-1} (Table 1, entry 2). This increase likely results from the additional stabilization induced into the primer/template duplex afforded by the formation of a s^{2}T:A base pair compared to a canonical T:A base pair,[16] as well as the more 3′-endo-like sugar puckering of 2-thio-nucleotide, which is the favoured sugar conformation in nonenzymatic primer extension reactions.[17] The k_{obs} values for reactions containing the activated ribonucleotides[18] 2-MeImpU, 2-MeImps^{2}U and its ribo-T analog 2-MeImps^{2}T were all

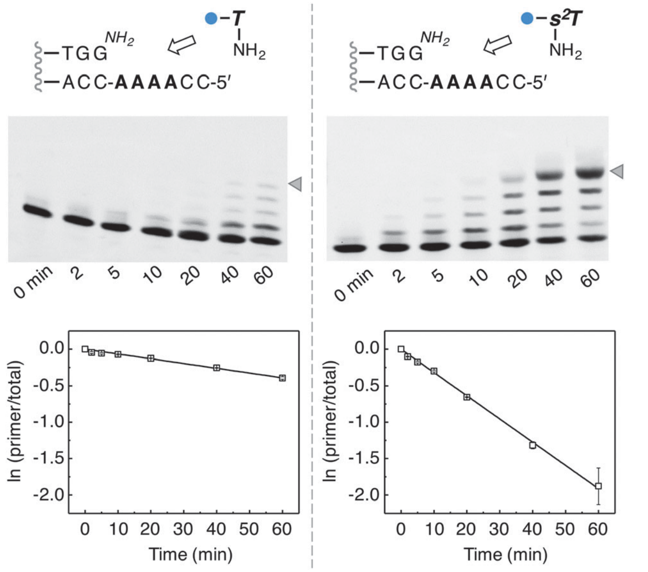


Fig. 2 Kinetics of copying a $r(A)_4(C)_2$ template with 3′-NH_2-2-MeImpddT (left) and 3′-NH_2-2-MeImpdds^{2}T (right), in the presence of 100 mM 1-(2-hydroxyethyl)-imidazole, at pH 7.5 and 4 °C. Reactions were initiated by addition of monomers and monitored by gel electrophoresis. The triangle indicates the primer +4 product. (bottom) Natural log of the fraction of the unreacted primer plotted against incubation time. Errors were based on two experiments. Primer strand (DNA, 0.2 μM): 5′-(FAM)-AGC-GTG-ACT-GAC-TGG-(NH_2)-3′, obtained enzymatically in *ca.* 85% purity based on LC-HRMS (see the ESI†). Primer concentration was corrected for unreactive oligonucleotide impurity. Template strand (RNA, 1 μM): 5′-CCAAAA-CCA-GUC-AGU-CAC-GCU-3′ RNA.

Table 1 Reaction kinetics measured for 10 mM T/U monomers at 4 °C

Entry	Template	Monomer	k_{obs} (h^{-1})	Relative k_{obs}
1[a]	$r(A)_4(C)_2$	3′-NH_2-2-MeImpddT	0.42 (1)	7
2[a]	$r(A)_4(C)_2$	3′-NH_2-2-MeImpdds^{2}T	1.92 (2)	30
3[b]	$r(A)_6$	2-MeImpU	ND	—
4[b]	$r(A)_6$	2-MeImps^{2}U	0.064 (1)	1
5[b]	$r(A)_6$	2-MeImps^{2}T	0.22 (6)	3

[a] In the presence of 100 mM 1-(2-hydroxyethyl)imidazole. [b] Data obtained from ref. 18. Reactions performed with 200 mM $MgCl_2$ at pH 7.0.

lower than the values for the activated 3′-amino nucleotides described above (Table 1), even though these ribonucleotide polymerizations were assayed in the presence of 200 mM Mg^{2+} to optimize the reactivity. The rate enhancement observed for 2-MeImps^{2}T *vs.* 2-MeImps^{2}U suggests that methylation at the 5-position of 2-thiouracil leads to stronger monomer–primer stacking. Additionally, primer extension reactions with 3′-NH_2-2-MeImpdds^{2}T are 10-fold faster than with the corresponding ribonucleotide, 2-MeImps^{2}T, presumably due to the greater nucleophilicity of the 3′-amine. Remarkably, combining the effect of the 3′-amine and the 5-methyl groups results in an *ca.* 30-fold (Table 1, entries 2 *vs.* 4) faster reaction. Further physical and kinetic characterization will be required to distinguish the contributions of enhanced monomer binding *vs.* enhanced monomer reactivity for these observations.

In conclusion, we have developed a scalable synthesis of a 2-thio modified thymidine monomer, 3′-NH_2-2-MeImpdds^{2}T. Our synthetic route provided this highly reactive nucleotide in sufficient amounts to perform quantitative measurements of nonenzymatic RNA template-copying rates for the first time. Our results show that 3′-NH_2-2-MeImpdds^{2}T can polymerize on a DNA/RNA primer/template complex significantly faster than any other U or T monomer that has been reported thus far.

We thank Dr Shao-Liang Zheng (Harvard) for help with X-ray data collection and structure determination, and Prof. Mohammad Movassaghi (MIT) and Dr Justin Kim (MIT) for their guidance in performing safe hydrogen sulfide experiments. We also thank Dr Lijun Zhou and Ms Moriana Haj for technical assistance and Dr Albert C. Fahrenbach, Dr Victor S. Lelyveld, Mr Tony Z. Jia, and Mr Chun Pong Tam for helpful discussions. J. W. S. is an Investigator of the Howard Hughes Medical Institute. This work was supported in part by a grant (290363) from the Simons Foundation to J. W. S.

Notes and references

1 M. P. Robertson and G. F. Joyce, The Origins of the RNA World, *CSHP Biol.*, 2012, **4**, a003608.
2 T. Inoue and L. E. Orgel, *J. Am. Chem. Soc.*, 1981, **103**, 7667.
3 W. S. Zielinski and L. E. Orgel, *Nucleic Acids Res.*, 1985, **13**, 2469.
4 M. Tohidi, W. S. Zielinski, C. H. Chen and L. E. Orgel, *J. Mol. Evol.*, 1987, **25**, 97.
5 V. Tereshko, S. Gryaznov and M. Egli, *J. Am. Chem. Soc.*, 1998, **120**, 269.
6 M. Kurz, K. Göbel, C. Hartel and M. W. Göbel, *Angew. Chem., Int. Ed. Engl.*, 1997, **36**, 842.
7 I. A. Kozlov, P. K. Politis, A. V. Aerschot, R. Busson, P. Herdewijn and L. E. Orgel, *J. Am. Chem. Soc.*, 1999, **121**, 2653.
8 S. Zhang, N. Zhang, J. C. Blain and J. W. Szostak, *J. Am. Chem. Soc.*, 2013, **135**, 924.
9 S. Zhang, J. C. Blain, D. Zielinska, S. M. Gryaznov and J. W. Szostak, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 17732.
10 Y. Masaki, R. Miyasaka, K. Hirai, H. Tsunoda, A. Ohkubo, K. Seio and M. Sekine, *Chem. Commun.*, 2012, **48**, 7313.
11 P. C. Tlatelpa and H. Huang, *Tetrahedron Lett.*, 2014, **55**, 4780–4784.
12 T.-S. Lin, Z.-Y. Shen, E. M. August, V. Brankovan, H. Yang, I. Ghazzouli and W. H. Prusoff, *J. Med. Chem.*, 1989, **32**, 1891.
13 This compound is 200 times cheaper (per 1 gram) than the starting material [5′-*O*-(dimethoxytrityl)-N^3/O^4-(toluoyl)-2-thiothymidine] that we used for the earlier synthetic route[9].
14 A. Miazga, F. Hamy, S. Louvel, T. Klimkait, Z. Pietrusiewicz, A. Kurzyñska-Kokorniak, M. Figlerowicz, P. Wiñska and T. Kulikowski, *Antiviral Res.*, 2011, **92**, 57.
15 E. Kervio, A. Hochgesand, U. E. Steiner and C. Richert, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 12074.
16 A. T. Larsen, A. C. Fahrenbach, J. Sheng, J. Pian and J. W. Szostak, *Nucleic Acids Res.*, 2015, **43**, 7675.
17 N. Zhang, S. Zhang and J. W. Szostak, *J. Am. Chem. Soc.*, 2012, **134**, 3691.
18 B. D. Heuberger, A. Pal, F. D. Frate, V. V. Topkar and J. W. Szostak, *J. Am. Chem. Soc.*, 2015, **137**, 2769.